



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/791,897

03/04/2004

Takahiko Kawatani

1509-477

4562

22879 7590 10/17/2007

HEWLETT PACKARD COMPANY
P O BOX 272400, 3404 E. HARMONY ROAD
INTELLECTUAL PROPERTY ADMINISTRATION
FORT COLLINS, CO 80527-2400

EXAMINER

PHAM, MICHAEL

ART UNIT

PAPER NUMBER

2167

MAIL DATE

DELIVERY MODE

10/17/2007

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/791,897

Applicant(s)

KAWATANI, TAKAHIKO

Examiner

Michael D. Pham

Art Unit

2167

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 05 July 2007.
- 2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-29 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-29 is/are rejected.
- 7) ☒ Claim(s) 7 and 10-12 is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

Art Unit: 2167

Detailed Action

Status of claims

1. Claims 1-29 are pending.

Priority

2. Should applicant desire to obtain the benefit of foreign priority under 35 U.S.C. 119(a)-(d) prior to declaration of an interference, a certified English translation of the foreign application must be submitted in reply to this action. 37 CFR 41.154(b) and 41.202(e).

Failure to provide a certified translation may result in no benefit being accorded for the non-English application.

Claim Objections

3. Claim 8 is objected to because of the following informalities: claim 8 recites "vale". It appears to be a typo. Appropriate correction is required. Examiner assumes the word "value".

Claim Rejections - 35 USC § 112

4. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5. Claim 1 -29 rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 1 recites the limitation "the initial state" and "the common co-occurrence matrix" in lines 9 and 14 of claim 1. There is insufficient antecedent basis for this limitation in the claim.

Claim 29 recites the limitation "the initial state" and "the common co-occurrence matrix" in lines 9 and 17 of claim 29. There is insufficient antecedent basis for this limitation in the claim.

Claim 9 contains a matrix T however, there is no description of what T is or does, and therefore claim 9 is indefinite.

Claim 7 contains "M equals the number of sorts of occurring terms Dr."; however, there is no M in (1). Therefore claim 7 is indefinite.

Claim 10 contains a matrix T and set D from a matrix Q however, there is no description of what T and set D from a matrix is or does, and therefore claim 10 is indefinite.

All other claims fail to resolve the deficiencies of the claims from which they depend.

Claim Rejections - 35 USC § 101

6. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

7. Claims 1, 13, and 29 under 35 U.S.C. 101 are respectfully withdrawn.

Claim Rejections - 35 USC § 103

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

Art Unit: 2167

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

9. Claims 1-6, 8, and 13-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent 7130848 by Oosta (hereafter Oosta) further in view of U.S. Patent Application 20050022106 by Kawai et. al. (hereafter Kawai).

Claim 1:

Oosta discloses the following claimed limitations:

“(a) obtaining a document or pattern frequency matrix for the set of input documents or patterns based on occurrence frequencies of terms appearing in each document or pattern;”[col. 10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number that represent frequency with which that word pair is found together in the all of the abstracts of the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line 57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set) based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-5, frequency with which that word pair is found together) is suggested]

“(C) obtaining the document or pattern commonality to the current cluster for each document or pattern in the input document or pattern set by using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on the common-co matrix of the current cluster, and making documents or

Art Unit: 2167

patterns having the document commonality higher than a threshold belong temporarily to the current cluster;” [col. 12 lines 24-30, the formation of a series of first technology topics composed of one or more words that are strongly related to each other. The collection of first technology topics is a second word matrix. Some words could be found in several first technology topics, and the common words define relationships between first technology topics. Accordingly, obtaining the document or pattern commonality to the current cluster (col. 12 lines 24-30, common words define relationships between first technology topics) for each document or pattern in the input document or pattern set by using information based on the document pattern frequency matrix for the input document or pattern (col. 11 lines 44-46, first technology topics be formed by associating high frequency word pairs from the first word correlation matrix), information based on the document or pattern frequency matrix for documents or patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from the first correlation matrix) and information based on the common-co matrix of the current cluster (col. 12 lines 30-34, second word matrix to further associate the related technology topics. The result is the formation of a set of second technology topics that are condensed versions of the first technology topics), and making documents or patterns having the document commonality higher than a threshold belong temporally to the current cluster (col. 12 lines 11-13, use of a threshold to form first technology topics can improve the focus of the first technology topics by illuminating stray words.) is suggested.]

“(d) repeating step (c)” [col. 12 lines 35-40, optionally further correlations can be conducted to form third, fourth, or fifth topics. Accordingly, (d) repeating step (c) (further correlations conducted) is suggested]

“(f) deciding, on the basis of the document or pattern commonality of each document or pattern to each cluster, a cluster to which each document or pattern belongs and outputting said cluster.”[col. 12 lines 53-56, assignment of a patent to a technology topic has been made based on the number of words from a technology topic that can be found in a patent abstract.

Accordingly, deciding (assignment), on the basis of the document pattern commonality of each document or pattern to each cluster (based on number of words), a cluster to which each documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11 line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose

“(b) selecting a seed document or pattern from remaining documents or patterns that are not included in any clustering existing at that moment and constructing a current cluster of the initial state using the seed document or pattern;”

“until the number of documents or patterns temporarily belong to the current cluster becomes the same as that in the previous repetition”

“(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and”.

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also

based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

On the other hand, Kawai discloses 0101 during the first phase, seed candidate documents 60 are evaluated to identify a set of seed documents 59. In 0103, stating only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents. In 0104, if the candidate seed documents being compared are not sufficiently distinct the candidate seed is grouped into a cluster 58 with the most similar cluster center 58 to which the candidate seed document was compared. Accordingly, until the number of documents or patterns temporarily belong to the current cluster (grouped into cluster 58) becomes the same as that in the previous repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are therefore within the same field of endeavor. For the above reasons, it would have been obvious to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of Oosta for the purpose of providing potential categories for clustering quickly, by using seed documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

“(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and”

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form technology topics. Accordingly, repeating step (c) until a given convergence condition is satisfied is (amount of identified word pair groups, the more topics that are formed, hence repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating steps (b) and (d) until a given convergence condition is satisfied (e.g. last candidate seed document is met)

Claim 2:

The combination of Oosta and Kawai disclose:

“(a-1) generating a document or pattern segment vector for each of said document or pattern segments based on occurrence frequencies of terms appearing in each document or pattern segment;” [Oosta, col. 10 lines 58, word correlation matrix]

“(a-2) obtaining a co-occurrence matrix for each document or pattern in the input document or pattern set from the document or pattern segment vectors; and”[Oosta, col. 12 lines 24-27, series of first technology topics composed of one or more words that are strongly related to each other. The collection of first technology topics is a second word matrix]

“(a-3) obtaining a document or pattern frequency matrix from the co-occurrence matrix for each document.” [Oosta, col. 11 lines 4-5, cell of the matrix contains a number that represent

the frequency with which that word pair is found together]

Claim 3:

The combination of Oosta and Kawai discloses:

“(b-1) constructing a common co-occurrence matrix of remaining documents or patterns that are not included in any cluster existing at that moment; and” [Kawai, 0047, linear vector representation and the scores are formed into normalized vectors.]

“(b-2) obtaining a document commonality to the set of the remaining document or pattern set for each document or pattern in the remaining document or pattern set by using the common co-occurrence matrix of the remaining documents or patterns, and extracting the document or pattern having the highest document or pattern commonality, and constructing a current cluster of the initial state by making a document or pattern set including the seed document or pattern and the neighbor documents or patterns similar to the seed document or pattern.” [Kawai, 0048, clustering module forms clusters of documents using the similarities of concepts and terms between the normalized vectors. As preparatory step in forming clusters, the clustering module iteratively analyzes a set of seed candidate documents to form a set of seed documents.]

Claim 4:

The combination of Oosta and Kawai discloses:

“(c-1) constructing a common co-occurrence matrix of the current cluster and a document or pattern frequency matrix of the current cluster;” [Oosta, col. 12 line 27, second word matrix]

“(c-2) obtaining the distinctiveness of each term and each term pair to the current cluster by comparing the document or pattern frequency matrix of the input document or pattern set and the document or pattern frequency matrix of the current cluster; and”[Oosta, col. 12 lines 32-34, the result is the formation of a set of secondary technology topics that are a condensed versions of the first technology topics. Col. 12 lines 11-13, use of threshold to form first technology topics can improve the focus of the first technology topics]

“(c-3) obtaining document or pattern commonalities to the current cluster for each document or pattern in the input document or pattern set by using the common co-occurrence matrix of the current cluster and weights of each term and term pair obtained from their distinctiveness, and making a document or pattern having the document or pattern commonality higher than a threshold belong temporarily to the current cluster.” [Oosta Col. 12 lines Col. 12 lines 11-13, use of threshold to form first technology topics can improve the focus of the first technology topics]

Claim 5:

The combination of Oosta and Kawai discloses:

“repeating step (e) until the number of documents or patterns whose document or pattern commonalities to any current clusters are less than a threshold becomes 0, or the number is less than a threshold and is equal to that of the previous repetition.”[Kawai, until next document is empty see, figure 14, element 176]

Claim 6:

The combination of Oosta and Kawai discloses:

“checking existence of a redundant cluster, and removing, when the redundant cluster exists, the redundant cluster and again deciding the cluster to which each document belongs.”

[Kawai, figure 14 element 166]

Claim 8:

The combination of Oosta and Kawai disclose:

“wherein each component of the document or pattern frequency matrix of a document or pattern set D is the number of documents or patterns in which a corresponding component of the co-occurrence matrix of each document or pattern in the document or pattern set D does not take a value of zero.” [Oosta, col. 11 lines 58-60, composed of all of the words in one column with a non-zero count.]

Claim 13:

The combination of Oosta and Kawai discloses:

“(a) obtaining a document or pattern commonality to the remaining document or pattern set for each document or pattern in the remaining document or pattern set by using the said common co-occurrence matrix of the remaining documents or patterns,” [Oosta, assignment of a patent to technology topic has been made based on the number of words from a technology topic that can be found in a patent abstract]

“(b) extracting, as candidates of the seed of the current cluster, a specific number of documents or patterns whose document or pattern commonalities obtained by step (a) are large;”

[Kawai, figure 14 element 161, identify candidate seed documents]

“(c) obtaining similarities of the respective candidates of the seed of the cluster to all documents or patterns in the input document or pattern set or in the remaining document or pattern set, and obtaining documents or patterns having similarities larger than a threshold as neighbor documents or patterns of the candidate; and” [Kawai, figure 14 element 168 and 167, group candidate seed documents into similar cluster]

“(d) selecting the candidate whose number of the neighbor documents or patterns is the largest among the candidates as the seed of the current cluster and making its neighbor documents or patterns the current cluster of the initial state.”[Kawai, figure 14 element 161, identify candidate seed documents]

Claim 14:

The combination of Oosta and Kawai discloses:

“detecting the distinctiveness of each term or object feature and each term pair with respect to the current cluster and detecting their weights,”[Kawai, 0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

the distinctiveness and weight detecting steps including

“(a) obtaining a ratio of each component of a document or pattern frequency matrix obtained from the input document or pattern set to a corresponding component of a document or pattern frequency matrix obtained from the current cluster as a document or pattern frequency ratio of each term or feature or each term or feature pair;”[Kawai, 0013, a frequency of occurrences of at least one concept within a document retrieved from the document set]

“(b) selecting a specific number of terms or features or term or feature pairs having the smallest document or pattern frequency ratios among a specific number of terms or features or term or feature pairs having the highest document or pattern frequencies, and obtaining the average of the document or pattern frequency ratios of the selected terms or features or term or feature pairs as the average document or pattern frequency ratio;”[0011, candidate seed documents evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents.]

“(c) dividing the average document or pattern frequency ratio by the document or pattern frequency ratio of each term or feature or each term or feature pair as a measure of the distinctiveness of each term or feature or each term or feature pair;”[0048, normalized vector]

“and (d) determining the weight of each term or feature or each term or feature pair from a function having the distinctiveness measure as a variable.”[0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

Claim 15:

The combination of Oosta and Kawai discloses:

“eliminating terms or features and term or feature pairs having document or pattern frequencies higher than a threshold.”[Oosta, col. 12 lines 4-6, a threshold can be set to accept word pairs into a first technology topic only if the count for that word is above the threshold]

Claim 16:

Art Unit: 2167

The combination of Oosta and Kawai discloses: "wherein clustering is performed recursively by letting the document or pattern set included in a cluster be the input document or pattern set."

[Figure 14 element 168, group candidate seed document into most similar cluster]

Claim 17:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 1" [Oosta col. 19 line 67, pc]..

Claim 18:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 2"[Oosta col. 19 line 67, pc]..

Claim 19:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 3"[Oosta col. 19 line 67, pc]..

Claim 20:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 4"[Oosta col. 19 line 67, pc]..

Claim 21:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 5" [Oosta col. 19 line 67, pc]..

Claim 22:

The combination of Oosta and Kawai discloses: "A computer program product for causing a computer to perform the method of claim 6" [Oosta col. 19 line 67, pc]..

Claim 23:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 1" [Oosta col. 19 line 67, pc].

Claim 24:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 2" [Oosta col. 19 line 67, pc]..

Claim 25:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 3"[Oosta col. 19 line 67, pc]..

Claim 26:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 4" [Oosta col. 19 line 67, pc]..

Claim 27:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 5" [Oosta col. 19 line 67, pc]..

Claim 28:

The combination of Oosta and Kawai discloses: "A computer arranged to perform the method of claim 6" [Oosta col. 19 line 67, pc]..

Claim 29:

Oosta discloses the following claimed limitations:

"A first unit for obtaining a document or pattern frequency matrix for the set of input documents or patterns, based on occurrence frequencies of terms appearing in each document or

pattern;” [col. 10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number that represent frequency with which that word pair is found together in the all of the abstracts of the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line 57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set) based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-5, frequency with which that word pair is found together) is suggested]

“a third unit for obtaining the document or pattern commonality to the current cluster for each document or pattern in the input document or pattern set using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on the common co-occurrence matrix of the current cluster and means for making documents or patterns having the document or pattern commonality higher than a threshold belong temporarily to the current cluster;” [col. 12 lines 24-30, the formation of a series of first technology topics composed of one or more words that are strongly related to each other. The collection of first technology topics is a second word matrix. Some words could be found in several first technology topics, and the common words define relationships between first technology topics. Accordingly, obtaining the document or pattern commonality to the current cluster (col. 12 lines 24-30, common words define relationships between first technology topics) for each document or pattern in the input document or pattern set by using information based on the document pattern frequency matrix for the input document or pattern (col. 11 lines 44-46, first technology topics be formed by associating high frequency word pairs from the first

Art Unit: 2167

word correlation matrix), information based on the document or pattern frequency matrix for documents or patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from the first correlation matrix) and information based on the common-co matrix of the current cluster (col. 12 lines 30-34, second word matrix to further associate the related technology topics. The result is the formation of a set of second technology topics that are condensed versions of the first technology topics), and making documents or patterns having the document commonality higher than a threshold belong temporally to the current cluster (col. 12 lines 11-13, use of a threshold to form first technology topics can improve the focus of the first technology topics by illuminating stray words.) is suggested.]

“a fourth unit for repeating the operations of the third unit” [col. 12 lines 35-40, optionally further correlations can be conducted to form third, fourth, or fifth topics. Accordingly, (d) repeating step (c) (further correlations conducted) is suggested]

“a sixth unit for deciding, on the basis of the document or pattern commonality of each document or pattern to each cluster, a cluster to which each document or pattern belongs, and for outputting said cluster.” [col. 12 lines 53-56, assignment of a patent to a technology topic has been made based on the number of words from a technology topic that can be found in a patent abstract. Accordingly, deciding (assignment), on the basis of the document pattern commonality of each document or pattern to each cluster (based on number of words), a cluster to which each documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11 line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose,

“second unit for selecting a seed document or pattern from remaining documents or patterns that are not included in any cluster existing at that moment and constructing a current cluster of the initial state using the seed document or pattern;”

“until the number of documents or patterns temporarily belonging to the current cluster becomes the same as that in the previous repetition;”

“a fifth unit for repeating the operations of the second through fourth units until given convergence conditions are satisfied; and”

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

On the other hand, Kawai discloses 0101 during the first phase, seed candidate documents 60 are evaluated to identify a set of seed documents 59. In 0103, stating only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents.

Art Unit: 2167

In 0104, if the candidate seed documents being compared are not sufficiently distinct the candidate seed is grouped into a cluster 58 with the most similar cluster center 58 to which the candidate seed document was compared. Accordingly, until the number of documents or patterns temporarily belong to the current cluster (grouped into cluster 58) becomes the same as that in the previous repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are therefore within the same field of endeavor. For the above reasons, it would have been obvious to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of Oosta for the purpose of providing potential categories for clustering quickly, by using seed documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

“a fifth unit for repeating the operations of the second through fourth units until given convergence condition is satisfied; and”

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form technology topics. Accordingly, repeating the third unit until a given convergence condition is satisfied is (amount of identified word pair groups, the more topics that are formed, hence repeats creation of topics until all identified word groups are made) suggested.

Art Unit: 2167

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating the second unit and fourth unit until a given convergence condition is satisfied (e.g. last candidate seed document is met)

Allowable Subject Matter

9. Claims 7 and 10-12 would be allowable if rewritten to overcome the rejection(s) under 35 U.S.C. 112, 2nd paragraph, set forth in this Office action and to include all of the limitations of the base claim and any intervening claims.

Response to Arguments

10. Applicant's arguments with respect to claim 7/5/07 and 7/17/07 have been considered but are moot in view of the new ground(s) of rejection.

Art Unit: 2167

Conclusion

11. The prior art made of record listed on PTO-892 and not relied, if any, upon is considered pertinent to applicant's disclosure.

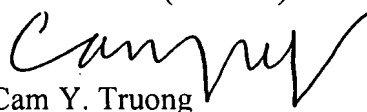
Contact Information

12. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael D. Pham whose telephone number is (571)272-3924. The examiner can normally be reached on Monday - Friday 9am - 5:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cottingham can be reached on 571-272-7079. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Michael Pham
Art Unit 2167
Examiner *M.P.*


Cam Y. Truong
Art Unit 2162
Primary Examiner

John Cottingham
Art Unit 2167
Supervisor